Parts and wholes – on the cognitive reality of multi-word units

Saskia E. Lensink, Niels O. Schiller and Arie Verhagen

According to usage-based theories of language, every time when we process or produce an utterance, our mental representation of the language changes (Bybee, 2006). This view would imply that when completely regular combinations of words occur frequently in the input, speakers are prone to store these combinations as wholes. Indeed, there is a growing body of psycho- and neurolinguistic studies that finds evidence for the existence of multi-word units in both language comprehension and language production (Arnon & Snider, 2010; Tremblay & Baayen, 2010; Tremblay & Tucker, 2011). Moreover, there is some evidence suggesting that even L2 speakers might store units larger than a word as a whole (Siyanova-Chanturia, Conklin & Van Heuven, 2011).

However, previous studies have only focused on English, and on a restricted set of multi-word units. In a series of recent experiments, we replicated and extended these findings for a wide range of multi-word units in Dutch and have found further evidence that L1 speakers and advanced L2 speakers make use of multi-word units.

We had Dutch participants read out loud frequent trigrams and quadgrams in their L1 or L2 (English). We extracted both onset latencies and durations and by using the newest mixed-effects modeling techniques (Wood, 2006), we found that the amount of information (Shannon, 1948) that multi-word units convey is more predictive of the speech rate in the first language of participants than the amount of information conveyed by the single words, although both seem to contribute. This suggests that participants make use of both a stored form of the whole multi-word unit, and the single words themselves. These findings on production data are reminiscent of predictions from dual route models where multiple processing routes are hypothesized to exist and can be employed at the same time in comprehension (Baayen, Dijkstra & Schreuder, 1997). Even more interestingly, in L2 speakers only the frequencies of the whole quadgrams seem to play a role in their speech rate, but the frequencies of the single words do not seem to significantly contribute to their speed of talking.

I will focus on the ways one can test for the cognitive reality of multi-word units, the optimal ways to analyze this type of data, and will discuss the implications of these findings for linguistic theory.

References


